

# A Simple Metric for Turn-Taking in Emergent Communication

5 August 2011

Running title: A Simple Metric for Turn-Taking

Authors:

Peter A. Raffensperger (Corresponding author)

Department of Electrical and Computer Engineering, University of Canterbury

Private Bag 4800

Christchurch, New Zealand

Phone: +64 3 364 2987 ext 7275

Fax: +64 3 364 2761

[peter.raffensperger@pg.canterbury.ac.nz](mailto:peter.raffensperger@pg.canterbury.ac.nz)

Russell Y. Webb

Apple Computer Inc.,

Cupertino, California, USA

Philip J. Bones

Department of Electrical and Computer Engineering, University of Canterbury,

Christchurch, New Zealand

[phil.bones@canterbury.ac.nz](mailto:phil.bones@canterbury.ac.nz)

Allan I. McInnes

Department of Electrical and Computer Engineering, University of Canterbury,

Christchurch, New Zealand

[allan.mcinnnes@canterbury.ac.nz](mailto:allan.mcinnnes@canterbury.ac.nz)

## **Abstract**

To facilitate further research in emergent turn-taking, we propose a metric for evaluating the extent to which agents take turns using a shared resource. Our measure reports a turn-taking value for a particular time and a particular time scale, or ‘resolution,’ in a way that matches intuition. We describe how to evaluate the results of simulations where turn-taking may or may not be present and analyse the apparent turn-taking that could be observed between random independent agents. We illustrate the use of our turn-taking metric by reinterpreting previous work on turn-taking in emergent communication and by analysing a recorded human conversation.

Keywords: Multi-agent systems, Decentralised systems, Resource allocation, Turn-taking

# 1 Introduction

‘Turn-taking’ can refer to a wide range of human social behaviours. Turn-taking in human speech has been studied extensively (Sacks et al., 1974; Wilson and Wilson, 2005; Stivers et al., 2009). In some contexts, turn-taking occurs in groups of machines (Di Paolo, 2000; Quinn, 2001; Iizuka and Ikegami, 2004) or groups comprised of people and machines (Turunen et al., 2006; Raux and Eskenazi, 2008). Consider two agents that share a resource that cannot be allocated to more than one agent at once. Each agent can try to use the resource at each of a number of time steps,  $t = 1, 2, 3, \dots$ . Suppose the history of agents trying to use the shared resource is as in Table 1.

Table 1: Two agents sharing a resource.

$t$	1	2	3	4
Agent 1	Tries to use resource	Does not try	Does not try	Tries
Agent 2	Does not try	Tries	Does not try	Tries

To what extent are agents 1 and 2 in Table 1 taking turns using the resource? Chao and Thomaz’s study of the interaction between an anthropomorphic robot and people lead them to a similar question: ‘What is an appropriate metric for good turn-taking?’ (Chao and Thomaz, 2010, p. 134) We propose to answer questions like this by introducing a metric for the quantity of turn-taking in groups of identical agents that share a single resource. We narrow our scope further by assuming a particular model of how the resource behaves when more than one agent tries to take a turn at one time. Our metric is a simple and intuitive one rather than a general one. However, we believe that our metric is useful in a variety of situations where it is possible for a group of agents to take turns perfectly, fail to take turns entirely, or have some behaviour that falls between those extremes. Situations like this occur in ‘emergent communication,’ where agents are learning to communicate. Agents learning to communicate may need to take turns to be successful but may display a range of other behaviours.

Frequently, groups of agents benefit from sharing information (Gmytrasiewicz and Durfee, 2001; Floreano et al., 2007) or at least having reliable access to common information (Amato et al., 2009). However, agents may not be innately endowed with the ability to communicate. Some agents can learn to communicate, either by learning a language common to the other agents (Grim et al., 2002; Smith, 2002; Goldman et al., 2007; Ampatzis et al., 2008) or by reusing their non-communication-specific abilities to create a communications channel (Quinn, 2001; Nolfi, 2005). Emergent communication refers to cases where the ability to communicate emerges in a group of adaptive agents that originally could not communicate. Wagner et al. (2003) and Nolfi (2005) survey previous research in emergent communication and discuss the conditions required for the emergence of communication.

Sometimes communication is over an interfering channel; turn-taking is one way of defeating interfer-

ence. People everywhere use turn-taking (Stivers et al., 2009) to defeat the difficulties in understanding another speaker while they themselves are talking. These difficulties relate to attention, short-term memory (Baddeley and Logie, 1999; Baddeley, 2000) and the limitations of language processing (Christofels, 2006). Furthermore, turn-taking shapes human languages through its role in conversation:

It seems productive to assume that, given conversation as a major, if not THE major, locus of a language's use, other aspects of language structure will be designed for conversational use and, *pari passu* [Latin: equally], for turn-taking contingencies (Sacks et al., 1974, p. 722).

The origins and characteristics of human language are the starting point of many emergent-communication studies (Cangelosi and Parisi, 2002). Realistic models of the emergence of communication should therefore include turn-taking.

Some emergent-communication studies assume that the group of agents already possesses the ability to coordinate turn-taking (Goldman et al., 2007, Definition 19). Other studies have specifically examined the emergence of turn-taking (Di Paolo, 2000; Quinn, 2001; Iizuka and Ikegami, 2004). One application of emergent turn-taking is medium access control for self-organising networks of heterogeneous wireless embedded systems. Sometimes researchers model interfering wireless networks as agents engaged in cooperative or competitive games (Larsson et al., 2009; Leshem and Zehavi, 2009; Yang et al., 2009). The optimal solution to some of those games is a kind of turn-taking where agents transmit one at a time (Larsson et al., 2009, p. 21). Our focus is on describing agents' behaviours rather than controlling those behaviours. A quantitative measure of turn-taking is required to facilitate future research in emergent turn-taking. Quantifying turn-taking facilitates algorithmic optimisation and machine learning of turn-taking behaviour.

Our contributions to the field of emergent communication are:

- We propose  $\tau\tau$ , a quantitative measure for turn-taking that integrates the concepts of fairness and efficiency. See Section 2.
- We derive the turn-taking performance of random agents in Section 3. We consider how to evaluate distributions and individual values of  $\tau\tau$  against the performance of random agents.
- We demonstrate the suitability of our turn-taking metric by re-interpreting a previously reported case of turn-taking in simulated agents in Section 4 (Di Paolo, 2000). We find that our computed values of  $\tau\tau$  are consistent with the conclusions drawn by Di Paolo. We present an application of our metric in Section 5 where we analyse the turn-taking behaviours present in a recorded human conversation.

We explore related work in Section 6. We discuss future work on alternative metrics and possible

extensions to  $\tau\tau$  in Section 7 and draw conclusions in Section 8.

## 2 A simple turn-taking metric, $\tau\tau(t, r)$

We propose a simple metric for turn-taking in groups of identical agents that share a limited resource. Our metric transforms the agents' actions into a single real number representing the quantity of turn-taking present. We consider the attempts of each agent to use the shared resource to compute a real-valued turn-taking allocation for each agent, given a particular range of times. The turn-taking metric,  $\tau\tau(t, r)$ , is the product of the fairness and efficiency of the agents' allocation. The parameters  $t$  and  $r$  represent the time that the turn-taking value refers to and the time scale, or 'resolution,' of the turn-taking measurement, respectively. Turn-taking behaviour may or may not be valuable to the agents. In general, our turn-taking metric does not inherently related to the agents' utilities but turn-taking is a valuable system behaviour in a number of emergent-communication studies (Di Paolo, 2000; Iizuka and Ikegami, 2004).

We assume all sequences are discrete in time and that all agents in the group simultaneously select actions to take for a particular time step,  $t$ . Some of those actions may involve trying to use the shared resource. We summarise agent  $a$ 's actions with a 'usage attempt sequence', defined as  $S_a(t) \in \{0, 1\}$ , where  $S_a(t) = 1$  indicates that agent  $a$  is attempting to use the shared resource at time  $t$ , and  $S_a(t) = 0$  indicates that agent  $a$  is not attempting to use the shared resource.

### 2.1 Resource allocation model

The resource allocation model provides a set of rules for allocating the shared resource. We need a rule, for example, if more than one agent tries to use the resource at a particular time (a collision). If none of the agents attempt to use the resource at a particular time, then the resource is unallocated. When only one agent makes a resource usage attempt, that agent gets the resource for that particular time. If all the agents could use the resource on each time step, then sharing the resource through turn-taking would be unnecessary. Resource collisions must give a lower value of the turn-taking measure than alternating single uses so as to match intuition. For simplicity, we stipulate that if more than one agent tries to use the resource at once, then no agent gets the resource. This way, the turn-taking of a usage attempt sequence where every time step has a collision is the same as a usage attempt sequence where no agent ever makes a usage attempt. Table 2 summarises our resource allocation model.

Each agent contributes to global turn-taking by taking its share of turns, that is, using the shared resource without collision on some time steps. The ratio of the number of time steps that an agent uses the shared resource without collisions to the resolution is its turn-taking 'allocation.' Within a set of  $A$  agents, we

Table 2: Resource allocation model summary.

Resource usage attempts	Allocations
No agent makes a resource usage attempt	No agent gets the resource
Only one agent makes a resource usage attempt	That agent gets the resource
More than one agent makes a usage attempt	No agent gets the resource

define agent  $a$ 's turn-taking allocation over the time range  $t$  to  $t + r - 1$  as:

$$\alpha_a(t, r) = \frac{1}{r} \sum_{s=t}^{t+r-1} \left( S_a(s) \prod_{b=1, b \neq a}^A [1 - S_b(s)] \right) \quad (1)$$

where  $r, t \in \mathbb{Z}$  and  $r > A$ . The parameter  $r$  is the resolution, the time scale for the turn-taking measurement. The expression  $S_a(s) \prod_{b=1, b \neq a}^A [1 - S_b(s)]$  takes a value of 1 if agent  $a$  and only agent  $a$  is attempting to use the shared resource at time  $s$  and equals 0 otherwise. Observe that  $\alpha_a(t, r) \in [0, 1]$ .

## 2.2 $\tau\tau$ is the product of fairness and usage efficiency

The allocation of the resource can only be 'fair' if all agents get access to it within a period of time corresponding to  $r$ . The resource usage fairness is therefore defined for a particular time,  $t$ , and a particular resolution,  $r$ , for  $A$  agents as:

$$fairness(t, r) = \begin{cases} 0 & \text{if } \sum_{a=1}^A \alpha_a(t, r) = 0 \\ \frac{A \min_a [\alpha_a(t, r)]}{\sum_{a=1}^A \alpha_a(t, r)} & \text{otherwise} \end{cases} \quad (2)$$

Thus we define fairness to be the normalised ratio of the allocation of the agent with the smallest allocation to the sum of the allocations for all agents, in the time range  $t$  to  $t + r - 1$ . For a truly fair situation,  $fairness(t, r) = 1$ ; for an unfair system, when one agent misses out altogether (say), fairness is zero. In general,  $fairness(t, r) \in [0, 1]$ .

This definition of fairness reflects our intuitive linkage between the human concepts of taking turns and of fairness. Jain's fairness index (Jain et al., 1984) is unsuitable in this case because it assigns non-zero fairness values in cases where one agents' allocation is zero. Other fairness metrics without the problem in Jain's metric are possible (Chen and Zhang, 2005; Lan et al., 2009). We discuss the possibility of using other fairness metrics in Section 8. Our chosen fairness metric matches our efficiency metric: the numerator of the efficiency metric is the same as the denominator of the fairness metric to simplify the expression for the turn-taking metric.

As well as being fair, our intuition says that turn-taking must be efficient. We define the usage efficiency to be the sum of the agents' allocations:

$$efficiency(t, r) = \sum_{a=1}^A \alpha_a(t, r) \quad (3)$$

In effect, the expression for  $efficiency(t, r)$  is the ratio of the number of steps with a single resource usage attempt to the total number of steps within the resolution. For example, if  $S_a(t) = 1$  for all  $a$  and  $t$ , then  $efficiency(t, r) = 0$ ; similarly if all  $S_a(t) = 0$  then  $efficiency(t, r) = 0$ . On the other hand, if  $S_1(t) = 1$  and  $S_a(t) = 0$  for  $a > 1$  and all  $t$ , then  $efficiency(t, r) = 1$ .

A different definition of efficiency could be more appropriate in some emergent-communication contexts. We could use the Pareto efficiency from economics (Musgrave and Musgrave, 1984; Brams, 2008), for example. We have chosen our definition to match our intuitive perception of turn-taking.

We define the degree of turn-taking,  $\tau\tau$ , as the product of the resource usage efficiency and the fairness:

$$\tau\tau(t, r) = efficiency(t, r) \times fairness(t, r) \quad (4)$$

$$= A \min_a [\alpha_a(t, r)] \quad (5)$$

where  $\tau\tau(t, r) \in [0, 1]$ . When  $\tau\tau(t, r) = 1$ , there is perfect turn-taking between all  $A$  agents in the time range  $t$  to  $t + r - 1$ ; in this case each agent makes the same number of resource usage attempts, no resource usage attempts collide and the agents use the resource on every time step. When  $\tau\tau(t, r) = 0$ , there is no turn-taking at all; in this case at least one agent has zero allocation. We can understand the values of  $\tau\tau(t, r)$  in terms of fairness and efficiency, as in Eq. (4), or in terms of the allocation of the agent with the smallest allocation, as in Eq. (5). For example, if  $\tau\tau(t, r) = 0.4$  then the agent with the fewest non-colliding resource usage attempts has 40% of the allocation it requires for perfect turn-taking.

### 2.3 Choosing a resolution, $r$

It is essential to choose the right resolution,  $r$ , when calculating  $\tau\tau(t, r)$ . Firstly, perfect turn-taking is impossible if the resolution is less than the number of agents: we must choose  $r \geq A$ . Equally, if the turn-taking exhibits a periodicity, then it would be reasonable to select  $r$  as equal to the period or an integer multiple of the period. In the periodic case,  $\tau\tau(t, r)$  is precise only for turn-taking periods that are perfect divisions of  $r$ . Considering an extreme case, suppose that  $A = 2$  with

$$S_1(t) = 1, 0, 1, 0 \text{ and}$$

$$S_2(t) = 0, 1, 0, 1 \text{ for } t = 1, 2, 3, 4.$$

In this case  $\tau\tau(1,4) = 1$ , but  $\tau\tau(1,3) = 2/3$ .  $efficiency(t,r) = 1$ , for all  $t$  and  $r$ , because exactly one agent attempts to use the resource at each time step.  $fairness(1,4) = 1$  because both agents get 2 turns each, but  $fairness(1,3) = 2\min[2,1]/(2+1) = 2/3$ . In this case, since the turn-taking has a period of 2,  $\tau\tau(t,3)$  is a less accurate measure of the turn-taking than  $\tau\tau(t,4)$ .

Consider perfect turn-taking with period  $T$ ,  $\tau\tau(t,T) = 1$  for all  $t$ . This implies  $\tau\tau(t,r) = 1$  for  $r = nT$ , where  $n \in \mathbb{Z}$  and  $n \geq 1$ . However,  $\tau\tau(t,r) = 1$  is not guaranteed for  $r \neq nT$ . As  $r$  increases, the accuracy of  $\tau\tau(t,r)$  increases because  $\tau\tau(t,r)$  averages over more periods of the turn-taking so ending alignment has less effect. Also, with larger  $r$ ,  $\tau\tau(t,r)$  will more accurately report turn-taking with longer periods. However, as  $r$  increases,  $\tau\tau(t,r)$  also becomes less specific to a particular time  $t$ . Thus we need to trade time localisation for accuracy.

### 3 Measuring the turn-taking of random agents

After performing a stochastic experiment with agents and measuring the turn-taking of the result, we have the question ‘What value of  $\tau\tau$  is sufficient to indicate that turn-taking behaviour is present rather than some other behaviour that, by chance, looks like turn-taking?’ We suggest that designed or learnt behaviour should do better than random actions, otherwise there is no value in the design or learning. Specifically, agents should significantly exceed the mean  $\tau\tau(t,r)$  of all possible probabilistic agents whose probabilities are independent of time and the other agents’ actions. We argue that it is sufficient to show that the agents being considered exceed the turn-taking of the probabilistic agents with the highest mean value of  $\tau\tau(t,r)$ , henceforth called the ‘best’ possible probabilistic agents. Specific values of  $\tau\tau(t,r)$  from an experiment should be compared against the distribution of  $\tau\tau(t,r)$  values from the best probabilistic agents.

For example, suppose there is a particular system that might display turn-taking for  $A = 3$  and  $r = 30$ . If the experimenter samples the turn-taking of the system for 100 independent trials and the mean turn-taking value is 0.3 with a standard deviation of 0.1, can the experimenter say that turn-taking has arisen? We can calculate the probability that such turn-taking occurred by chance if we know the distribution of  $\tau\tau(t,30)$  for groups of 3 random agents with the best possible usage attempt probabilities. As we will calculate formally below, for  $r = 30$  and  $A = 3$ , the distribution has a mean of 0.27 and a standard deviation of 0.12. Our statistical null hypothesis is that the mean  $\tau\tau(t,30)$  of the system is less than or equal to the mean  $\tau\tau(t,30)$  of the random agents. We test this hypothesis using a one-sided  $t$ -test with unequal variances and sample sizes (Snedecor and Cochran, 1989). The  $t$ -test assumes normal distributions, but is robust to violations of this assumption when the sample sizes are large, as is the case here. We get a one-sided  $p$ -value of 0.03 indicating that there is a 3% chance that the system actually has a lower mean  $\tau\tau(t,30)$  value than the random agents. So, for 95% confidence level, we should reject the null hypothesis, and accept the alternative hypothesis that the system has a larger mean turn-taking



than we could expect due to chance.

Suppose that we have a group of probabilistic agents, where we know the probability of agent  $a$  attempting to use the shared resource,  $p[S_a(t) = 1]$ , for all  $A$  agents. Let these probabilities be independent of the other agents' actions and of time, so  $p[S_a(t) = 1] = P_a$ . We view the turn-taking value of these probabilistic agents as a random variable, denoted  $TT(A, r, P_1, P_2, \dots, P_A)$ . The distribution of  $TT(A, r, P_1, P_2, \dots, P_A)$  depends on the resolution,  $r$ , the number of agents,  $A$ , and the probabilities,  $P_a$ ,  $1 \leq a \leq A$ . We can calculate this distribution from all possible usage attempt sequences at some resolution, or estimate it from a random sample of usage attempt sequences.

We define the best possible probabilistic agents as the ones with the largest expected turn-taking value,  $E[TT(A, r, P_1, P_2, \dots, P_A)]$ . Because  $\tau\tau(t, r)$  assumes that all agents are identical, a unique maximum for  $E[TT(A, r, P_1, P_2, \dots, P_A)]$  will occur only when  $P_a = P$  for  $1 \leq a \leq A$ . Based on the resource allocation model of Eq. (1), each agent's average allocation is maximised if each agent takes an average of one turn per  $A$  time steps. Therefore, the maximum of  $E[TT(A, r, P_1, P_2, \dots, P_A)]$  is when  $P = 1/A$ . If  $P_a = 1/A$  for all agents, then  $TT(A, r, P_1, P_2, \dots, P_A)$  is a random variable representing the turn-taking of the best possible agents and will be henceforth referred to as  $TT(A, r)_{best}$ . Table 3 shows estimates of the means and standard deviations of  $TT(A, r)_{best}$  for various numbers of agents at resolutions that are various multiples of  $A$ , estimated from  $10^6$  samples of different possible usage attempt sequences.

Table 3: The mean,  $E(TT_{best})$ , and standard deviation,  $\sigma(TT_{best})$ , turn-taking values for the best probabilistic agents for various resolutions,  $r$ , and numbers of agents,  $A$ . These values are estimates computed from a population of  $10^6$  samples of the possible usage attempt sequences.

$A$		$r = A$	$r = 2A$	$r = 10A$	$r = 100A$	$r \rightarrow \infty$
2	$E(TT_{best})$	0.13	0.23	0.37	0.46	0.5
	$\sigma(TT_{best})$	0.33	0.27	0.14	0.045	0
3	$E(TT_{best})$	0.019	0.095	0.27	0.39	0.44
	$\sigma(TT_{best})$	0.14	0.20	0.12	0.041	0
4	$E(TT_{best})$	0.0029	0.042	0.22	0.36	0.42
	$\sigma(TT_{best})$	0.054	0.14	0.11	0.038	0
5	$E(TT_{best})$	0.00044	0.018	0.19	0.34	0.41
	$\sigma(TT_{best})$	0.021	0.094	0.098	0.036	0
10	$E(TT_{best})$	0.00	0.00034	0.12	0.29	0.39
	$\sigma(TT_{best})$	0.00	0.013	0.077	0.031	0
100	$E(TT_{best})$	0.00	0.00	0.0080	0.23	0.37
	$\sigma(TT_{best})$	0.00	0.00	0.027	0.022	0

We can evaluate particular values of  $\tau\tau(t, r)$  for some  $A$  and  $r$  against the distribution of  $TT(A, r)_{best}$ . If a group of agents takes turns with some value of  $\tau\tau(t, r)$ , then the probability that random agents would not achieve at least that same value,  $p[TT(A, r)_{best} < \tau\tau(t, r)]$ , is the probability that the group of agents is taking turns by some means other than chance. However, for very small resolutions, the large variance

in the distribution of  $TT(A, r)_{best}$  means that only low confidence levels can be achieved. If there is a population of turn-taking values, one can compare that population with the distribution of  $TT(A, r)_{best}$ . We can evaluate turn-taking at any resolution if we have a large enough sample of turn-taking values, but using single values of  $\tau\tau(t, r)$  is only meaningful with large resolutions.

We now derive the  $\lim_{r \rightarrow \infty} E[TT(A, r)_{best}]$  as a function of the number of agents,  $A$ . The expected allocation for random agents as  $r \rightarrow \infty$  is:

$$\lim_{r \rightarrow \infty} E[\alpha_a(t, r)] = P \prod_{b=1, b \neq a}^A (1 - P) \quad (6)$$

$$= P(1 - P)^{A-1} \quad (7)$$

The best random agents have usage attempt probabilities  $P = 1/A$ :

$$\lim_{r \rightarrow \infty} E[\alpha_a(t, r)_{best}] = \frac{1}{A}(1 - 1/A)^{A-1} \quad (8)$$

The expected value of  $TT(A, r)_{best}$  can be calculated by Eq. (5) for the limit as  $r \rightarrow \infty$ :

$$\lim_{r \rightarrow \infty} E[TT(A, r)_{best}] = A \min_a \left( \lim_{r \rightarrow \infty} E[\alpha_a(t, r)_{best}] \right) \quad (9)$$

$$= A \left( \frac{1}{A} (1 - 1/A)^{A-1} \right) \quad (10)$$

$$= \left( \frac{A-1}{A} \right)^{A-1} \quad (11)$$

Eq. (11) is at its maximum of 0.5 at  $A = 2$  and monotonically decreases with increasing  $A$ , approaching  $1/e$  in the limit:

$$\lim_{A \rightarrow \infty} \lim_{r \rightarrow \infty} E[TT(2, r)_{best}] \rightarrow 1/e \approx 0.37 \quad (12)$$

The trend towards this limit is evident in Table 3.

## 4 Application of the metric to work by Di Paolo

Di Paolo (2000) simulated pairs of mobile agents that used acoustic signalling to approach each other. Each agent had an acoustic sensor and a motor on each half of its round body, and a single acoustic transmitter. Recurrent neural networks controlled the agents. Di Paolo examined a number of aspects of his simulation. We only expand on the interpretation of the turn-taking behaviour of his agents' acoustic signalling; we add nothing to his other results. Each of Di Paolo's agents could detect the other agent's position based on the other's simulated acoustic signalling. Di Paolo found that the agents would activate their transmitters alternately at a constant frequency and that this alternating transmission helped them move better. The acoustic channel was additive and limited in range; one agent's signalling interfered

with the other's signalling.

#### 4.1 Defining the resource usage attempt sequences

For our purposes, Di Paolo’s agents take turns using the shared resource of the simulated acoustic channel. We focus on Fig. 13 of Di Paolo’s paper that shows the alternating signalling of his two agents. We extracted the numerical values from Di Paolo’s graph to make Fig. 1. The graphical extraction process introduces small errors but does not significantly affect our results. The domain of Di Paolo’s original graph is 200 units of time. We have quantised the domain to 2000 time steps across the graph, 10 time steps for each of Di Paolo’s units. Henceforth, we abandon Di Paolo’s original indexing of time that ranges between 400 and 600, and refer only to our new indices that range between 0 and 2000. The signalling in Fig. 1 occurs after the agents’ learning has converged to a solution to their coordinated motion task.

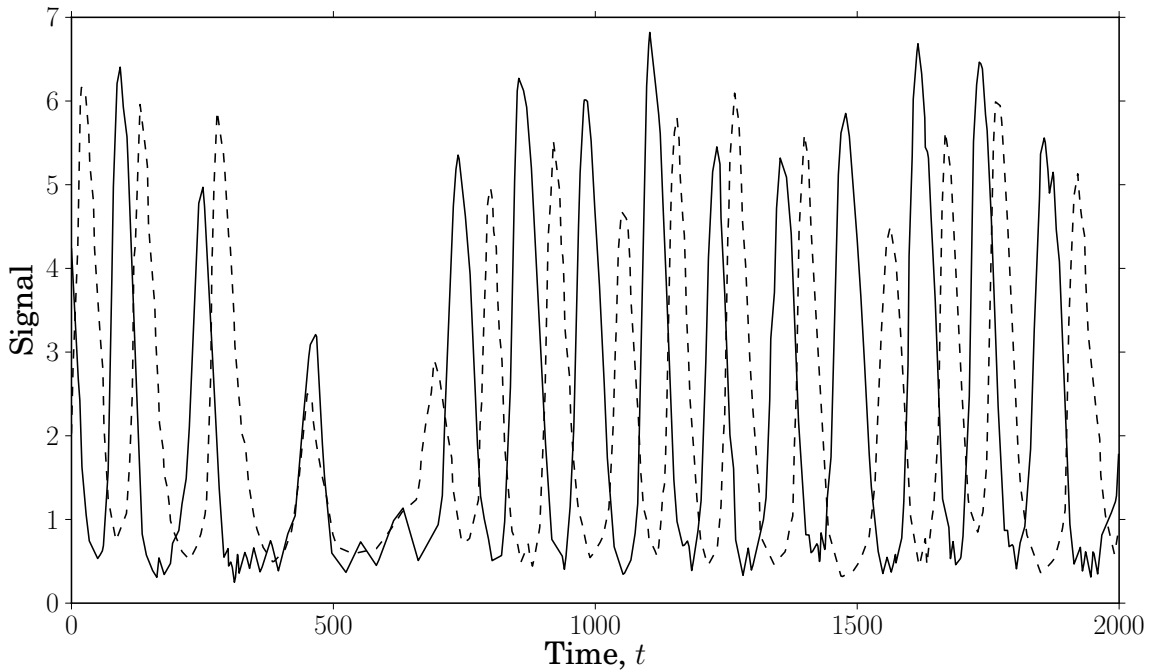


Figure 1: Turn-taking emerged in a simulation by Di Paolo. The solid line represents one agent’s signalling and the dashed line represents the other’s. We extracted the data for this graph from Fig. 13 of Di Paolo (2000).

The turn-taking values depend strongly on how we define the usage attempt sequences; Di Paolo does not specify a way of defining the turn boundaries. Because we cannot assign specific meanings to signal features, we arbitrarily define the usage attempt sequences by thresholding the values in Fig. 1: for agent  $a$ ,  $S_a(t) = 1$  if  $Signal_a(t) > 2$  and  $S_a(t) = 0$  otherwise. Setting the threshold higher would result in fewer collisions but also less productive use of the channel. Fig. 2 shows the resulting resource

allocation attempt sequences  $S_1$  and  $S_2$  plotted against time. Using a threshold is arbitrary and the value of the threshold is arbitrary, but the resulting usage attempt sequences serve to illustrate our turn-taking metric. Di Paolo observes that “Since agents have no way of knowing of the presence of the other but through acoustic coupling an efficient way of doing this [completing their coordinated motion task] is by alternating the production of signals and so minimizing overlap.” (Di Paolo, 2000). Fig. 3 shows that the usage attempt sequences have some collisions; perhaps Di Paolo’s agents try to minimise collisions but still some collisions are present.

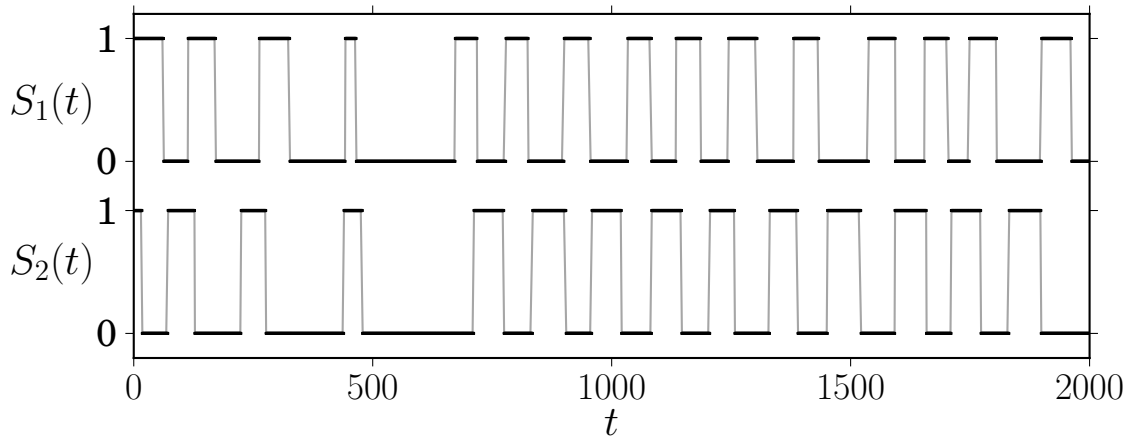


Figure 2: Two agents from Di Paolo (2000) share an acoustic resource. Their resource usage attempt sequences,  $S_1$  and  $S_2$ , are plotted against time using data from Fig. 1.

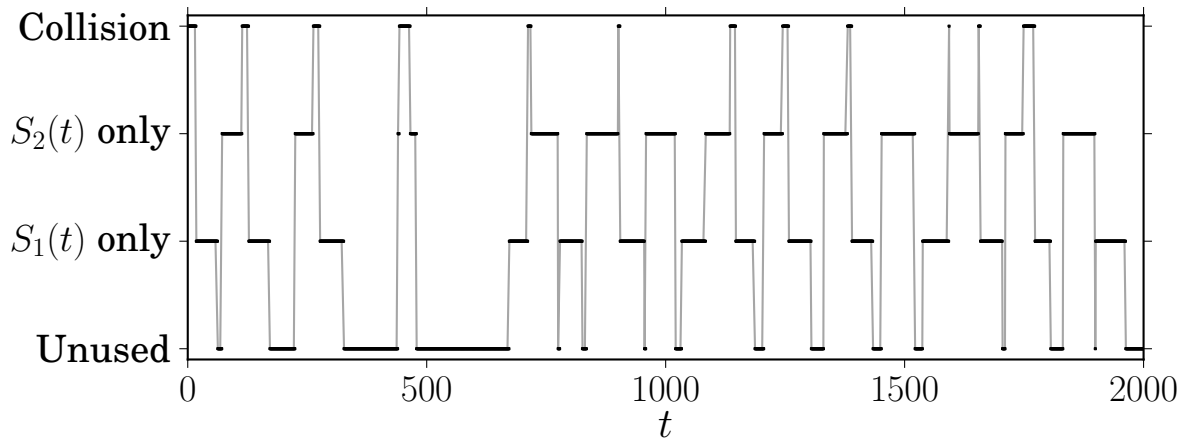


Figure 3: Analysis of the resource usage attempt sequences,  $S_1$  and  $S_2$ , from Fig. 2. Collisions occur when  $S_1(t) = S_2(t) = 1$  and the channel remains unused when  $S_1(t) = S_2(t) = 0$ . Much of the time, however, only one of  $S_1(t)$  or  $S_2(t)$  equals 1.

## 4.2 Choosing a resolution

Fig. 3 shows that the two agents alternate their usage attempts much of the time, with small gaps and overlaps in between. Intuitively, we expect some turn-taking to be present. However, the resource is mostly unused in the range of time around  $327 < t < 672$ , aside from a collision around  $t = 450$ ; this time range should get low turn-taking values. Our choice of resolution,  $r$ , should be small enough to discriminate this low turn-taking time range from the rest of the sequence.

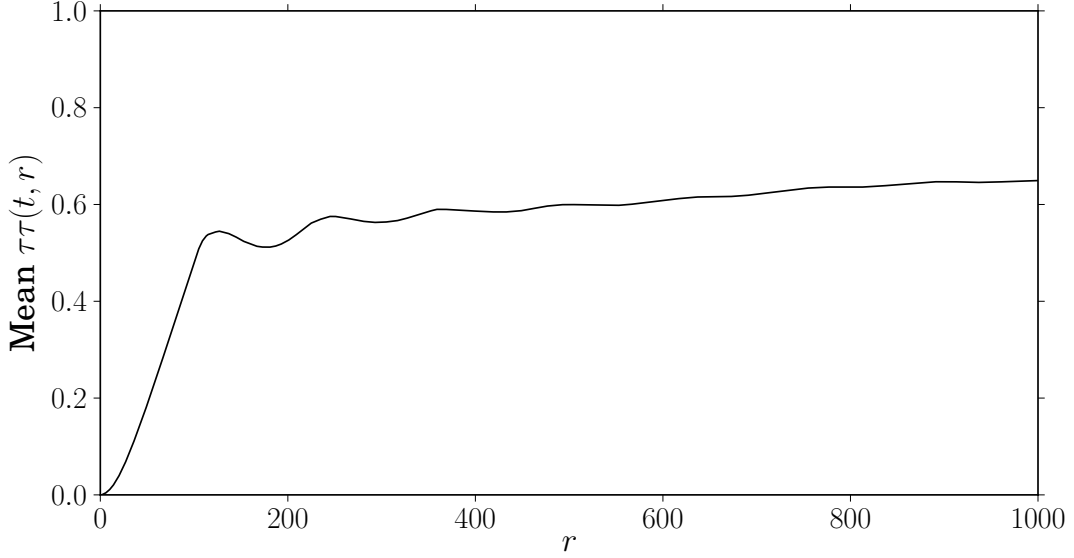


Figure 4: Mean turn-taking over time,  $\tau\tau(t, r)$ , generally increases with resolution,  $r$ , but has ripples with local maxima occurring near integer multiples of the period of turn-taking in the usage attempt sequences. The positions of these local maxima can be used to estimate the period of turn-taking.

The second half of Fig. 2 has 8 pulses of  $S_1$  and 7 pulses of  $S_2$ , giving about  $7\frac{1}{2}$  periods of turn-taking in 1000 time steps. This suggests that  $r = 133$  would be reasonable. Fig. 4 shows the mean turn-taking over time as a function of the resolution and illustrates the effects of changing the resolution described in Section 2.3:  $\tau\tau(t, r)$  mostly increases with increasing  $r$ , but has ripples with local maxima at  $r \approx r_0, 2r_0, 3r_0, \dots$  and local minima at  $r \approx 1.5r_0, 2.5r_0, 3.5r_0, \dots$ , where  $r_0$  is an estimate of the period of turn-taking present in  $S_1$  and  $S_2$ . By inspection,  $r_0 = 125$ , closely matching our choice of  $r = 133$ .

## 4.3 Results

Fig. 5 shows the variation of  $\tau\tau(t, 133)$  over time. The domain of Fig. 5 does not extend to the last 133 time steps; calculating  $\tau\tau$  in the last 133 time steps would require values for the usage attempt sequences for times with  $t > 2000$ . Because we compute  $\tau\tau(t, r)$  from the values of  $S_a(t)$  from  $t$  to  $t + r - 1$ ,  $\tau\tau(t, r)$  refers to the turn-taking in the next  $r$  steps. The low turn-taking range at  $327 < t < 672$  is correctly identified;  $\tau\tau(t, 133) < 0.2$  for  $250 < t < 599$ . Because  $\tau\tau(t, r)$  refers to the next  $r$  steps, it

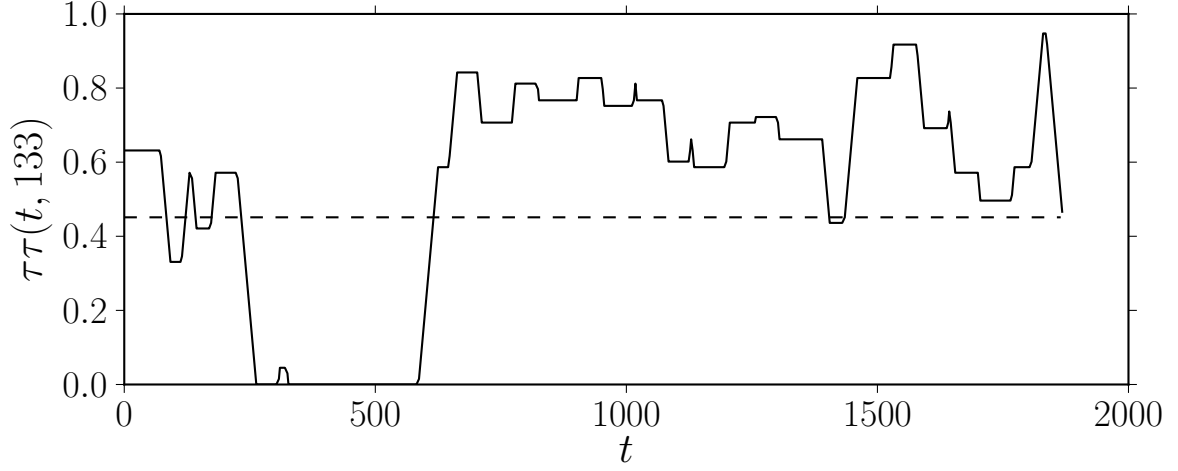


Figure 5: Turn-taking with  $r = 133$ ,  $\tau\tau(t, 133)$ , varies with time. A period of low turn-taking is present at  $277 < t < 580$ . The expected turn-taking with a resolution of 133 for pairs of random agents,  $E[TT(2, 133)_{best}]$ , has a value of 0.45 and is plotted as a horizontal dashed line.

makes sense that the low turn-taking time range occurs earlier in  $\tau\tau(t, r)$  than it does in  $S_1(t)$  and  $S_2(t)$ . After  $t = 599$ , most of the graph exceeds 0.45, the maximum expected turn-taking at a resolution of 133 for pairs random agents.

Table 4: Mean and standard deviation over time for  $\tau\tau$ , efficiency and fairness at  $r = 133$ .

	Mean	Standard deviation
$\tau\tau(t, 133)$	0.54	0.29
$efficiency(t, 133)$	0.66	0.28
$fairness(t, 133)$	0.70	0.33

From Table 4, we can see that turn-taking is present on average over the entire sequence. Using the methods described in Section 3, we estimated the distribution of  $TT(2, 133)_{best}$  with  $10^6$  samples. The probability that random agents will achieve a turn-taking value of at least 0.54 is 0.031. Therefore, for a confidence of 95%, turn-taking at a resolution of 133 is present on average over the sequence.

The large standard deviation of  $\tau\tau(t, 133)$ , 0.29, is due to the period of low turn-taking time range in the first half of the sequence; the mean of  $\tau\tau(t, 133)$  in the second half of the sequence is 0.67, with a standard deviation of 0.12. In the time range where turn-taking does not seem to be present,  $275 < t < 575$ , the mean turn-taking is 0.003 indicating no turn-taking with probability  $> 99\%$  as compared to random agents.

Di Paolo observed the anti-phase locking of the signalling in Fig. 1 and made the qualitative assessment

that turn-taking was present in the signalling. He also showed that the internal activations of his agents’ neural networks were strongly negatively correlated in periods of high qualitative turn-taking and only weakly negatively correlated in periods of low qualitative turn-taking (Di Paolo, 2000). Our turn-taking measure  $\tau\tau(t, r)$  matches Di Paolo’s qualitative judgement.

## 5 Application of the metric to a human conversation

We now apply the metric to the analysis of a recorded human conversation and we discuss how our quantitative results highlight social effects that may be otherwise overlooked. The behaviour of Di Paolo’s simulated agents in Section 4 was amenable to intuitive analysis in his original paper and our quantitative results support his conclusions. On the other hand, a recorded conversation is more difficult to analyse: extracting the usage attempt sequence is more complicated, the duration of each turn varies greatly and turns do not occur regularly. Furthermore, the metric assumes that all agents are identical, as described in Section 2, but people are all different. While we expect turn-taking to be present in conversation in general (Sacks et al., 1974; Stivers et al., 2009), the research on turn-taking in sociology gives us no specific insights to the particular conversation we analyse here.

We measure the quantity of turn-taking present in a one minute excerpt from a corpus of speech called ‘CONversational Speech In Noisy Environments’ (COSINE). The COSINE corpus was created by Stupakov et al. (2009) to assist studies on automatic speech recognition and related topics. The corpus includes 33 conversation sessions with two to seven participants per session. Each participant wore an array of seven microphones and participated in natural conversations in noisy environments. Stupakov et al. (2009) and Stupakov et al. (2012) describe the data collection and transcription. We analyse a conversation from Session 2, starting 7 minutes 47.5 seconds into the session and ending at 8 minutes 47.5 seconds. Our source data are the speech recordings from the close-talking microphone on each person’s array, which has the least noise and the best isolation between speakers. Four speakers, two male and two female young adults, introduce themselves to each other and converse in fluent English.

### 5.1 Defining the resource usage attempt sequences

As in Section 4, we must extract usage attempt sequences from continuous-valued signals. We want the usage attempt sequence to represent the times when each speaker is talking. Transforming a speech signal into a binary valued signal indicating when the speaker is talking is the well-studied problem of voice activity detection; for one example algorithm, see Sohn et al. (1999). We use a simpler approach: each audio signal is processed with a one-pole high pass filter with a cut-off frequency of 478 Hz followed by an envelope follower. When the envelope of the audio signal exceeds a hand selected threshold, we deem that agent to be speaking. We hand select a different threshold for each speaker to

provide a balance between rejecting noise and failing to recognise speech. The resulting usage attempt sequences have some noise, but have sufficient quality for our purposes. The original audio is at 44100 Hz, but we down-sample the usage attempt sequences to 100 Hz on the assumption that a person will talk for at least 10ms at time and also to decrease the required computational time. Fig. 6 shows the waveforms of the four speakers after the high pass filter preprocessing stage and the resulting usage attempt sequences. Analysing the turn-taking in Fig. 6 would be difficult without a turn-taking metric.

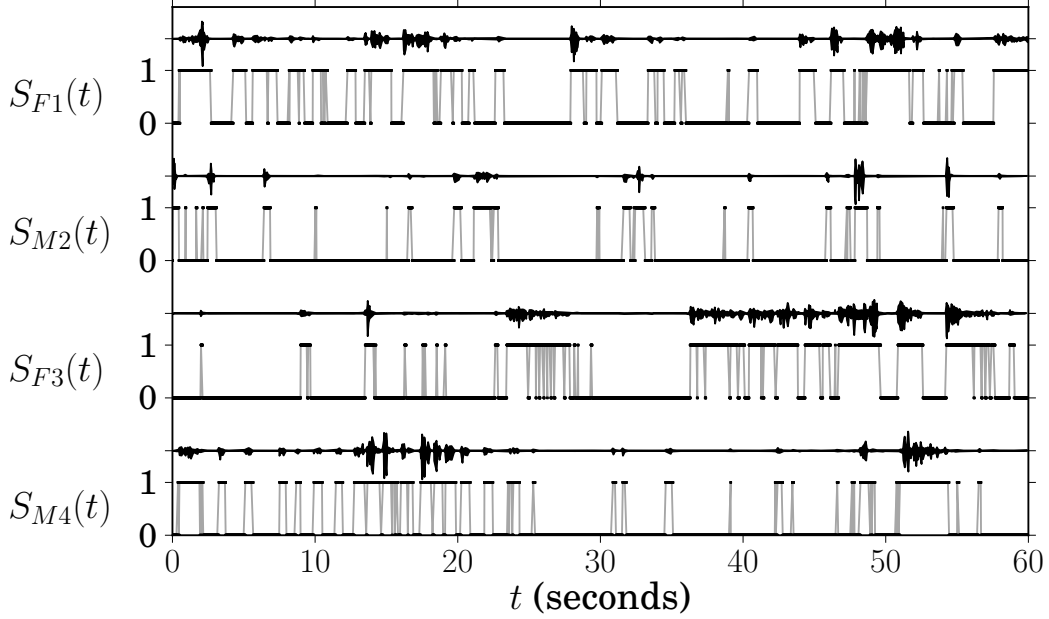


Figure 6: Sound amplitude waveforms and the extracted usage attempt sequences for the four speakers for the chosen excerpt from the COSINE corpus (Stupakov et al., 2009).

## 5.2 Choosing a resolution

Fig. 6 shows that the usage attempt sequences do not form an obvious pattern. However, agent F3 talks while the other agents are not talking for a time between about 22.5 seconds and 27.5 seconds. We expect low turn-taking for this time period while it is possible that turn-taking is present at other times. In order to discriminate this 5 second period of no turn-taking, we must use a resolution of no more than 5 seconds. On the other hand, smaller windows suffer from sensitivity to ending alignment as discussed in Section 2.3. With this balance in mind, we chose a resolution of 500 ten-millisecond time steps for a total window size of 5 seconds.



### 5.3 Results

Fig. 7 shows the calculated turn-taking at a resolution of 5 seconds as a function of time. Only on two occasions does the turn-taking of the four people exceed  $E[TT(4, 500)_{best}]$ , the expected turn-taking of groups of four random agents. We might find the overall low quantity of turn-taking surprising for a human conversation. Sacks et al. (1974) found that some form of turn-taking is a usual feature of conversation and indeed the four speakers in our conversation follow Sacks et al.’s turn trade rules (Sacks et al., 1974, p. 704). However, Sacks et al. observed that in unstructured conversation, ‘Relative distribution of turns is not specified in advance’ (Sacks et al., 1974, p. 701), whereas our turn-taking metric requires a fair distribution of speaking time between each agent. The lack of fair and efficient turn-taking makes sense if we ignore the content of the conversation and focus solely on who is talking when. The conversation has a total of 6.83 seconds when no one is speaking, while a single person is speaking for a total of 33.87 seconds and at least two people are speaking at once for a total of 19.3 seconds. The mean efficiency at a resolution of 5 seconds is 0.57. However, the dominant contribution to the low turn-taking values is fairness: the mean fairness at a resolution of 5 seconds is only 0.045. We consider the possibility of alternative turn-taking metrics that use alternative fairness metrics in Section 7.

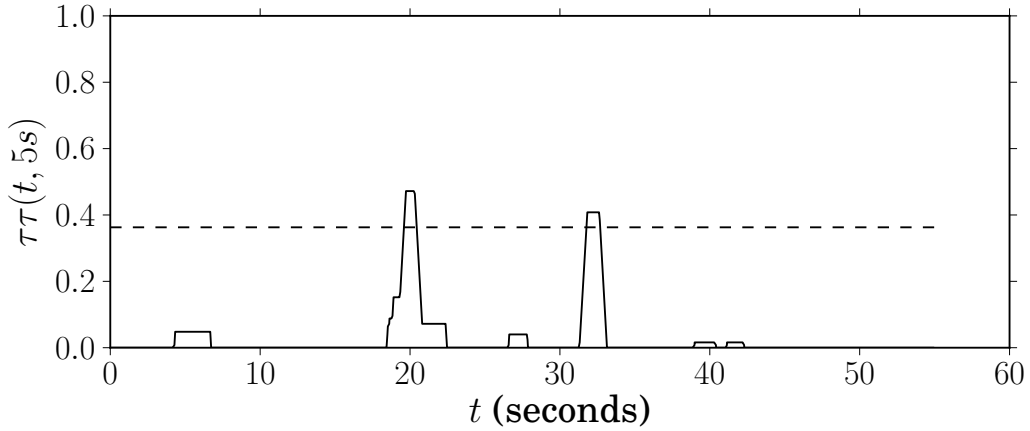


Figure 7: Turn-taking at a resolution of 5 seconds (500 time steps of 10ms) plotted against time for a 60-second conversation between four young adults. The expected turn-taking at a resolution of 500 for groups of four random agents,  $E[TT(4, 500)_{best}]$ , has a value of 0.36 and is plotted as a horizontal dashed line.

The results of our turn-taking metric allow us to quantify social effects. The low average value of the fairness metric reflects the disparity between agents’ time speaking. Our metric assumes that all the agents are identical (see Section 2). The four participants in the study are socially equal: their 2008 American culture has no significant gender biases in this context, they are similar ages (19, 19, 20 and 22), and they all received the same instructions from the researchers. The differences in the agents stem from other factors, such as personality or interest in the topics of conversation. Our turn-taking metric highlights differences like these and can motivate further examination. Agent M2 speaks

significantly less than the other agents, as shown in Fig. 6. The only two short periods where  $\tau\tau(t, 5s) > E[TT(4, 500)_{best}]$  in Fig. 7 coincide with the times that agent M2 speaks without another agent speaking over him (see Fig. 6).

Agents F1 and M4 dominate the first half of the conversation and agents F1 and F3 dominate the second half of the conversation. If we consider the conversation to be between agents F1 and M4 only, then there is considerably more turn-taking at a resolution of 5 seconds at most times, as shown in Fig. 8. The quantity of turn-taking for agents F1 and M4 is still generally below  $E[TT(2, 500)_{best}]$ ; this is unsurprising since there are many times when the agents under consideration (F1 and M2) stop speaking to listen to the agents M2 and F3.

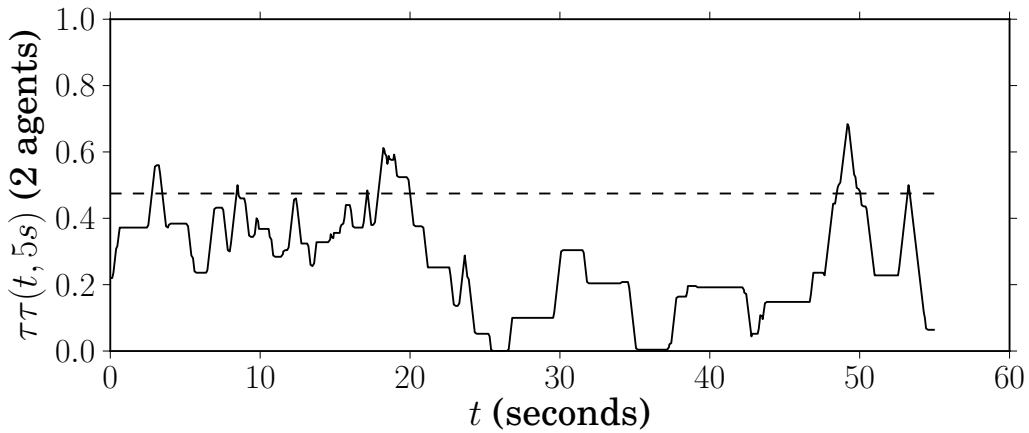


Figure 8: Turn-taking at a resolution of 5 seconds (500 time steps of 10ms) plotted against time, for a human conversation considering only agents F1 and M4. The expected turn-taking at a resolution of 500 for pairs of random agents,  $E[TT(2, 500)_{best}]$ , has a value of 0.47 and is plotted as a horizontal dashed line.

## 6 Related work

The properties of our turn-taking metric are inspired by work in a number of existing fields, including emergent communication. Previously, authors measured emergent turn-taking intuitively, or with *ad hoc* methods such as qualitative examination of a graph or by finding negative correlation between two signals with similar power spectra (Di Paolo, 2000). Our turn-taking metric is quantitative and is only sensitive to turn-taking, unlike a correlation.

Iizuka and Ikegami (2004) simulated pairs of agents that moved in a uniform, flat space and took turns chasing each other. Each agent had a circular body with a motor on either side. A recurrent neural network controlled the motors. A genetic algorithm selected the weights in the agents' neural networks; an agent's fitness depended on how well it predicted its partner's motion and how well it took turns chasing its partner. We compare  $\tau\tau(t, r)$  in the context of Iizuka and Ikegami's experiment with their

own turn-taking metric,  $F^{turn}$ :

$$\tau\tau(1, r) = \frac{2}{r} \min_a \left( \sum_{t=1}^r S_1(t) [1 - S_2(t)], \sum_{t=1}^r S_2(t) [1 - S_1(t)] \right) \quad (13)$$

$$F^{turn} = \alpha_1 \alpha_2 \quad (14)$$

where  $\alpha_1 = \sum_{t=1}^r S_1(t)$  and  $\alpha_2 = \sum_{t=1}^r S_2(t)$ , and  $S_1(t) = 1$  if and only if the position of agent  $a$  is in a circular sector behind the other agent and  $S_1(t) = 0$  otherwise (Iizuka and Ikegami, 2004, Eqs. 8 and 9). (Iizuka and Ikegami use different notation, with  $g_a(t)$  and  $g_b(t)$  in place of  $S_1(t)$  and  $S_2(t)$  and with  $T$  in place of  $r$ .) One key difference between the two metrics is that  $F^{turn}$  uses a resource allocation model that does not allow for the possibility of both agents attempting to use the shared resource at the same time (Iizuka and Ikegami, 2004, p. 3). If  $S_1(t) = 1 \iff S_2(t) = 0$  and  $S_2(t) = 1 \iff S_1(t) = 0$  then  $\tau\tau(1, r) = \frac{2}{r} \min(\alpha_1, \alpha_2)$ . Another fundamental difference is that large  $\tau\tau(t, r)$  values require fairness through the  $\min(\alpha_1, \alpha_2)$  term for all values of  $\alpha_1$  and  $\alpha_2$ .  $F^{turn}$  does not require fairness in the same way. The maximum of  $F^{turn}$  is when  $\alpha_1 = \alpha_2 = r/2$ , which is a fair allocation between the two agents. However, at other points in its range,  $F^{turn}$  increases with both  $\alpha_1$  and  $\alpha_2$ ;  $dF^{turn}/d\alpha_1 = \alpha_2 \geq 0$  and  $dF^{turn}/d\alpha_2 = \alpha_1 \geq 0$ . This means that highly unfair allocations can produce large  $F^{turn}$  values. The gradient of  $F^{turn}$  could be a strength in the context of Iizuka and Ikegami (2004) if it helps the genetic algorithm find the maximum but we consider that requiring fairness is essential for a turn-taking metric. Iizuka and Ikegami do not specify a way to generalise their metric to more than two agents.

Neill (2003) describes the ‘turn-taking dilemma,’ a variant on the prisoners’ dilemma (Flood, 1958) where the optimal strategy is for the players to alternate between defecting and cooperating. This situation appears in the natural behaviour of groups of dwarf mongoose (Anne and Rasa, 1989). Neill uses a concept of ‘evolutionary dominance’ (Neill, 2001) to rank strategies with memories of length 1 and 2 in the presence of noise. Neill evaluates strategies that may or may not take turns rather than evaluating histories of actions as we have done in this paper. Because a turn-taking dilemma rewards turn-taking most highly, the best strategies are good at coordinating turn-taking with other good strategies. ‘Since mutual cooperation and mutual defection result in suboptimal payoffs, a failure to coordinate one player’s cooperation with the other player’s defection results in harm to both players’ (Neill, 2003, p. 243). Neill’s concept of ‘self-alternating’ strategies relates to our intuition that turn-taking should be fair. However, the best of Neill’s strategies exploit more naïve strategies, resulting in worse turn-taking. ‘Since each player is expected to defect half the time, exploitation occurs when one player takes more than his fair share of turns, harming the other player.’ (Neill, 2003, p. 243). Action records with high payoffs in a turn-taking dilemma game will also have a large value of  $\tau\tau(t, r)$ .

Lan et al. (2009) derive a family of fairness functions based on a set of axioms. Some of Lan et al.’s metrics also incorporate a notion of efficiency. The definition of fairness in Eq. (2) is related to a Lan et

al. fairness measure,  $f_\beta$ , where  $\beta \rightarrow \infty$ . We can write Eq. (2) as

$$fairness(t, r) = \frac{A}{-f_\beta(\alpha_a(t, r), 1 \leq a \leq A)} \quad (15)$$

where  $f_\beta(\mathbf{x}) = -\max_i\{(\sum_i x_i)/x_i\}$  for  $\beta \rightarrow \infty$ . Lan et al. consider unbounded fairness functions so the transformation of Eq. (15) is necessary to limit the range of the fairness measure to  $[0, 1]$ . Using Lan et al.'s Eq. 5 for the weights would mean that the axiom of irrelevance of partition does not apply to our fairness measure. However, our fairness metric satisfies the continuity, homogeneity, asymptotic saturation, and monotonicity axioms. While our fairness metric bears similarity to one of Lan et al.'s metrics, Lan et al. make no attempt to measure turn-taking specifically or to consider the fair allocation of a resource through time.

We have considered the fair and efficient allocation of a limited shared resource through time in the context of emergent communication. Many researchers have explored efficient, fair resource allocation in general. Brams (2008) examines a number of fair division problems from a political science perspective. Bouveret and Lang (2008) investigate the computational complexity of finding efficient allocations of indivisible resources that satisfy an 'envy-free' notion of fairness. However, neither Brams nor Bouveret and Lang considers the distribution of allocations through time as we have with our turn-taking measure.

Researchers have also devised ways of measuring the quality of turn-taking in human-human or human-machine speech. Turn-taking is present in conversation when speaker changes occur and the participants follow turn trading rules (Sacks et al., 1974). Measuring turn-taking is important in spoken dialog systems, such as automated telephone answering agents, because it contributes to user satisfaction. In spoken dialog systems, turn-taking quality depends on the lengths of silences between speech and the interruption rates (Raux and Eskenazi, 2008; Turunen et al., 2006). Measuring silences and interruptions is related to our efficiency metric which forms part of our turn-taking metric. Efficient interaction is also important in human-robot interactions. Chao and Thomaz (2010) examine turn-taking between an anthropomorphic robot and a person as the person teaches the robot to sort coloured objects; Chao and Thomaz observe that good turn-taking has 'minimal over-lapping of turns as well as minimal time spent between turns, characteristics of an efficient interaction' (Chao and Thomaz, 2010, p. 133). The key difference between human-human or human-machine turn-taking measures and our work is that we do not assume the presence of any kind of turn-taking a priori. Counting silences and interruptions is a good measure of deviations from normal human-generated turn-taking but it fails to distinguish monologues from conversations because fairness is not considered. Our metric classifies a wider range of possible agent interactions, making it better suited for use in emergent communication.

## 7 Discussion

Observe that  $\tau\tau(t, r)$  is not the only possible turn-taking metric and one could make other decisions than those we have made for  $\tau\tau(t, r)$ . In the following, we consider the limitations in applicability of  $\tau\tau(t, r)$  and alternative decisions that would lead to other turn-taking metrics.

We assume that all agents are identical and unchanging with time. However, this will not always be the case. It is possible that agents may need the shared resource for different amounts of time, for example if one agent asks another a series of long questions with short answers. We could scale each agent's allocation in Eq. (1) to account for differences in agents' needs. Our choice of fairness metric only depends on the agent with the smallest allocation, but sometimes it is only important that some of the agents are participating in turn-taking. For example, if three people are in a room and two are conversing while the other is silent, how much turn-taking is present? Our metric resolves this question by focusing on the person who is silent: they take no turns, so the group has a turn-taking value of zero. For situations where we desire a different answer to that question, a different fairness metric would be appropriate.

Our resource allocation model is simplistic; others are definitely possible. The essential properties of a turn-taking resource allocation model are that leaving the resource unused results in low allocation and allocating to more than one agent at a time gives a lower allocation than allocating to a single agent at a time. In general, the cost and possibility of collisions will vary from situation to situation and the turn-taking metric may need to be modified. For example, the cost of a collision could be a function of the number of agents colliding at that time step, so that a collision between 100 out of 100 agents is worse than a collision between 2 of 100 agents.

We assume binary usage attempts, but many situations involve continuous usage attempts. Alternative turn-taking metrics could represent agents trying to use only part of the shared resource, with  $S_a(t)$  taking any real value between 0 and 1, rather than  $S_a(t)$  taking either 0 or 1. Another possibility is to approximate a continuous resource with a binary usage attempt sequence, as in Section 4.1 where we use a threshold. We could have indexed  $\alpha_a(t, r)$ ,  $\text{fairness}(t, r)$ ,  $\text{efficiency}(t, r)$  and  $\tau\tau(t, r)$  as depending on  $S_a(s)$  values for  $t - r + 1 \leq s \leq t$ . However, our choice of indexing from  $t$  to  $t + r - 1$  allows an allocation value to be calculated at  $t = 0$  for finite usage attempt sequences starting at  $t = 0$  without having to assume values of  $S_a(s)$  for  $t < 0$ .

In general, turn-taking metrics must compute resource allocations given usage attempts, reflect the concepts of fairness and efficiency, and deal appropriately with issues of resolution. Our decisions leading to  $\tau\tau(t, r)$  serve our goal of creating a simple metric for turn-taking but limit the scope of our metric's applicability. This limitation is intentional: while our methodology for measuring turn-taking is more general, we opt to open the discussion on measuring turn-taking with a specific, simple metric for groups of identical agents. We leave the possible extensions to  $\tau\tau(t, r)$  and alternative metrics to future work.

## 8 Conclusion

We propose  $\tau\tau(t, r)$ , a quantitative metric for the degree of turn-taking present in resource usage attempt sequences by identical agents. This metric allows evaluation of turn-taking in emergent-communication experiments. We define  $\tau\tau(t, r)$  as the product of the allocation efficiency and fairness to reflect our intuitive understanding of turn-taking. The concept of resolution is essential to measuring turn-taking: turn-taking may be present on some time scales but not others. We suggest that intentional turn-taking behaviour should have a value of  $\tau\tau(t, r)$  that exceeds the maximum expected turn-taking that would be produced by random agents. We give means and standard deviations for the distributions of  $\tau\tau(t, r)$  produced by random agents and methods for comparing observed values and distributions of  $\tau\tau(t, r)$  with those distributions from random agents. We show that our metric is useful in practice by applying it a simulation by Di Paolo (2000) and to a recorded human conversation taken from the COSINE corpus Stupakov et al. (2009). Our new metric enables us to make quantitative measurements about the presence of turn-taking rather than informal qualitative judgements.

## **Acknowledgements**

Thanks to Hiroyuki Iizuka (Human Information Eng. Lab., Information Science and Technology, Osaka University) and Takashi Ikegami (Graduate School of Arts and Sciences, The University of Tokyo) for data that came from their experiments (Iizuka and Ikegami, 2004) which was useful in developing and validating our metric. The authors acknowledge also helpful comments by the reviewers of the first manuscript submitted. The work of Peter Raffensperger is supported in part by the NZi3 ICT Innovation Institute and by a Canterbury Doctoral Scholarship from the University of Canterbury.

## References

- Amato, C., Bernstein, D. S., and Zilberstein, S. (2009). Optimizing fixed-size stochastic controllers for POMDPs and decentralized POMDPs. *Autonomous Agents and Multi-Agent Systems*, 21(3):293–320.
- Ampatzis, C., Tuci, E., Trianni, V., and Dorigo, M. (2008). Evolution of signaling in a multi-robot system: Categorization and communication. *Adaptive Behavior*, 16(1):5.
- Anne, O. and Rasa, E. (1989). The costs and effectiveness of vigilance behaviour in the dwarf mongoose: implications for fitness and optimal group size. *Ethology Ecology & Evolution*, 1(3):265–282.
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, 4(11):417–423.
- Baddeley, A. D. and Logie, H. R. (1999). Working memory: The multiple-component model. In Miyake, A. and Shah, P., editors, *Models of working memory: Mechanisms of active maintenance and executive control*, pages 28–61. Cambridge University Press, Cambridge, UK.
- Bouveret, S. and Lang, J. (2008). Efficiency and envy-freeness in fair division of indivisible goods: Logical representation and complexity. *Journal of Artificial Intelligence Research*, 32(1):525–564.
- Brams, S. J. (2008). *Mathematics and Democracy: Designing Better Voting and Fair-Division Procedures*. Princeton University Press, Princeton, N.J.
- Cangelosi, A. and Parisi, D. (2002). *Simulating the Evolution of Language*. Springer-Verlag New York, Inc. New York, NY, USA, London.
- Chao, C. and Thomaz, A. L. (2010). Turn taking for human-robot interaction. In *Dialog with Robots: Papers from the AAAI Fall Symposium*. Association for the Advancement of Artificial Intelligence.
- Chen, Z. and Zhang, C. (2005). A new measurement for network sharing fairness. *Computers & Mathematics with Applications*, 50(5-6):803–808.
- Christoffels, I. (2006). Listening while talking: The retention of prose under articulatory suppression in relation to simultaneous interpreting. *European Journal of Cognitive Psychology*, 18(2):206–220.
- Di Paolo, E. A. (2000). Behavioral coordination, structural congruence and entrainment in a simulation of acoustically coupled agents. *Adaptive Behavior*, 8(1):27–48.
- Flood, M. M. (1958). Some experimental games. *Management Science*, 5(1):5–26.
- Floreano, D., Mitri, S., Magnenat, S., and Keller, L. (2007). Evolutionary conditions for the emergence of communication in robots. *Current Biology*, 17(6):514–519.
- Gmytrasiewicz, P. J. and Durfee, E. H. (2001). Rational communication in Multi-Agent environments. *Autonomous Agents and Multi-Agent Systems*, 4(3):233–272.



- Goldman, C., Allen, M., and Zilberstein, S. (2007). Learning to communicate in a decentralized environment. *Autonomous Agents and Multi-Agent Systems*, 15(1):47–90.
- Grim, P., Denis, P. S., and Kokalis, T. (2002). Learning to communicate: The emergence of signaling in spatialized arrays of neural nets. *Adaptive Behavior*, 10(1):45.
- Iizuka, H. and Ikegami, T. (2004). Adaptability and diversity in simulated turn-taking behavior. *Artificial Life*, 10(4):361–378.
- Jain, R., Chiu, D., and Hawe, W. (1984). A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. Technical report, Digital Equipment Corporation.
- Lan, T., Kao, D., Chiang, M., and Sabharwal, A. (2009). An axiomatic theory of fairness. Technical report, Princeton University.
- Larsson, E. G., Jorswieck, E., Lindblom, J., and Mochaourab, R. (2009). Game theory and the flat fading gaussian interference channel. *IEEE Signal Processing Magazine*.
- Leshem, A. and Zehavi, E. (2009). Game theory and the frequency selective interference channel. *IEEE Signal Processing Magazine*.
- Musgrave, R. A. and Musgrave, P. B. (1984). *Public Finance in Theory and Practice*. McGraw-Hill, New York, 4th edition.
- Neill, D. B. (2001). Optimality under noise: Higher memory strategies for the alternating prisoner’s dilemma. *Journal of Theoretical Biology*, 211(2):159–180.
- Neill, D. B. (2003). Cooperation and coordination in the turn-taking dilemma. In *Proceedings of the 9th conference on Theoretical aspects of rationality and knowledge*, pages 231–244, Indiana, USA. ACM.
- Nolfi, S. (2005). Emergence of communication in embodied agents: co-adapting communicative and non-communicative behaviours. *Connection Science*, 17(3):231–248.
- Quinn, M. (2001). Evolving communication without dedicated communication channels. In Kelemen, J. and Sosik, P., editors, *Advances in Artificial Life: Sixth European Conference*, volume 2159 of *Lecture Notes in Computer Science*, pages 357–366.
- Raux, A. and Eskenazi, M. (2008). Optimizing endpointing thresholds using dialogue features in a spoken dialogue system. In *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, pages 1–10.
- Sacks, H., Schegloff, E. A., and Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.
- Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science*, 14(1):65.

- Snedecor, G. W. and Cochran, W. G. (1989). *Statistical Methods*. Iowa State University Press, 8 edition.
- Sohn, J., Kim, N. S., and Sung, W. (1999). A statistical model-based voice activity detection. *Signal Processing Letters, IEEE*, 6(1):1–3.
- Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., Ruitter, J. P. D., Yoon, K. E., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26):10587.
- Stupakov, A., Hanusa, E., Bilmes, J., and Fox, D. (2009). COSINE - a corpus of multi-party CONversational Speech In Noisy Environments. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pages 4153 –4156.
- Stupakov, A., Hanusa, E., Vijaywargi, D., Fox, D., and Bilmes, J. (2012). The design and collection of COSINE, a multi-microphone in situ speech corpus recorded in noisy environments. *Computer Speech & Language*, 26(1):52–66.
- Turunen, M., Hakulinen, J., and Kainulainen, A. (2006). Evaluation of a spoken dialogue system with usability tests and long-term pilot studies: similarities and differences. In *Ninth International Conference on Spoken Language Processing*, pages 1057–1060.
- Wagner, K., Reggia, J. A., Uriagereka, J., and Wilkinson, G. S. (2003). Progress in the simulation of emergent communication and language. *Adaptive Behavior*, 11(1):37.
- Wilson, M. and Wilson, T. P. (2005). An oscillator model of the timing of turn-taking. *Psychonomic bulletin & review*, 12(6):957.
- Yang, J., Klein, A. G., and Brown, D. R. (2009). Natural cooperation in wireless networks. *Signal Processing Magazine, IEEE*, 26(5):98 –106.